Chapitre 3

Principales distributions de probabilités

Introduction

De nombreuses situations pratiques peuvent être modélisées à l'aide de variables aléatoires qui sont régies par des lois spécifiques. Il importe donc d'étudier ces modèles probabilistes qui pourront nous permettre par la suite d'analyser les fluctuations de certains phénomènes en évaluant, par exemple, les probabilités que tel événement ou tel résultat soit observé.

La connaissance de ces lois théoriques possède plusieurs avantages sur le plan pratique :

- Les observations d'un phénomène particulier peuvent être remplacées par l'expression analytique de la loi où figure un nombre restreint de paramètres (1 ou 2, rarement plus).
- La loi théorique agit comme modèle (idéalisation) et permet ainsi de réduire les irrégularités de la distribution empirique. Ces irrégularités sont souvent inexplicables et proviennent de fluctuations d'échantillonnage, d'imprécision d'appareils de mesure ou de tout autre facteur incontrôlé ou incontrôlable.
- Des tables de probabilités ont été élaborées pour les lois les plus importantes. Elles simplifient considérablement les calculs.

Ce cours présente trois distributions discrètes : la distribution binomiale, la distribution géométrique et la distribution de Poisson. Puis il aborde deux distributions continues : la distribution exponentielle et la distribution normale. Il importe de bien comprendre quelles sont les situations concrètes que l'on peut modéliser à l'aide de ces distributions. Viennent enfin trois distributions théoriques dont la fonction n'est pas de modéliser mais de servir d'outils dans les problèmes d'estimation et de test.

3.1 Distribution binomiale

3.1.1 Variable de Bernoulli ou variable indicatrice

Définition

Définition 1 *Une variable aléatoire discrète qui ne prend que les valeurs 1 et 0 avec les probabilités respectives p et $q = 1 - p$ est appelée variable de Bernoulli.*

Exemple 2 *Une urne contient deux boules rouges et trois boules vertes. On tire une boule de l'urne. La variable aléatoire X = nombre de boules rouges tirées est une variable de Bernoulli. On a : $p(X = 1) = 2/5 = p$, $p(X = 0) = 3/5 = q$.*

Plus généralement, on utilisera une variable de Bernoulli lorsqu'on effectue une épreuve qui n'a que deux issues : le succès ou l'échec. Une telle expérience est alors appelée épreuve de Bernoulli. On affecte alors 1 à la variable en cas de succès et 0 en cas d'échec.

Distribution de probabilités

x	0	1
$f(x) = p(X = x)$	q	p

Paramètres de la distribution

On calcule

$$\begin{aligned}
E(X) &= 0.q + 1.p = p, \\
V(X) &= E(X^2) - E(X)^2 = (0^2 q + 1^2 p) - p^2 = p - p^2 = pq, \\
\xi_X(u) &= E(e^{-2i\pi uX}) = 1.q + e^{-2i\pi u}p = q + p\cos(2\pi u) + ip\sin(2\pi u).
\end{aligned}$$

$E(X) = p$	$V(X) = pq$	$\sigma(X) = \sqrt{pq}$	$\xi_X(u) = q + pe^{-2i\pi u}$

3.1.2 Distribution binomiale

Situation concrète

a) On effectue une épreuve de Bernoulli. Elle n'a donc que deux issues : le succès avec une probabilité p ou l'échec avec une probabilité q.

b) On répète n fois cette épreuve.

c) Les n épreuves sont indépendantes entre elles, ce qui signifie que la probabilité de réalisation de l'événement "succès" est la même à chaque épreuve et est toujours égale à p.

Dans cette situation, on s'intéresse à la variable X = "nombre de succès au cours des n épreuves".

Distribution de probabilités

Appelons X_i les variables de Bernoulli associées à chaque épreuve. Si la i-ème épreuve donne un succès, X_i vaut 1. Dans le cas contraire X_i vaut 0. La somme de ces variables comptabilise donc le nombre de succès au cours des n épreuves. On a donc $X = X_1 + X_2 + \cdots + X_n$. X peut prendre $n + 1$ valeurs : $0, 1, \ldots, n$.

Cherchons la probabilité d'obtenir k succès, c'est-à-dire $p(X = k)$.

La probabilité d'avoir k succès suivis de $n - k$ échecs est $p^k q^{n-k}$ car ces résultats sont indépendants les uns des autres.

La probabilité d'avoir k succès et $n - k$ échecs dans un autre ordre de réalisation est toujours $p^k q^{n-k}$. Donc tous les événements élémentaires qui composent l'événement $(X = k)$ ont même probabilité.

Combien y en a-t-il ' Autant que de façons d'ordonner les k succès par rapport aux $n - k$ échecs ' Il suffit de choisir les k places des succès parmi les n possibles et les $n - k$ échecs prendront les places restantes. Or il y a $\binom{k}{n}$ manières de choisir k places parmi n.

Finalement, on obtient

$$p(X = k) = \binom{k}{n} p^k q^{n-k}.$$

On dit que la variable aléatoire X suit une *loi binomiale de paramètres n et p*. On note $X \hookrightarrow B(n, p)$.

Remarque : L'adjectif binomial vient du fait que lorsqu'on somme toutes ces probabilités, on retrouve le développement du binôme de Newton,

$$\sum_{k=0}^{n} \binom{k}{n} p^k q^{n-k} = (p + q)^n = 1.$$

NB : La loi binomiale est tabulée en fonction des 2 paramètres n et p.

Paramètres descriptifs de la distribution

Nous savons que $X = X_1 + \cdots + X_n$ avec $E(X_i) = p$ pour $i = 1, 2, \ldots, n$, donc $E(X) = E(X_1) + \cdots + E(X_n) = np$.

Les variables X_i sont indépendantes et $Var(X_i) = pq$ pour $i = 1, 2, \ldots, n$, donc $Var(X) = Var(X_1) + \cdots + Var(X_n) = npq$. D'autre part, les fonctions caractéristiques se multiplient, donc $\xi_X(u) = (q + pe^{-2i\pi u})^n$.

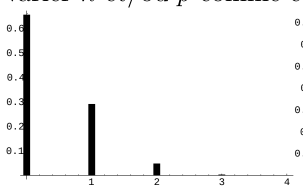
$E(X) = np$	$V(X) = npq$	$\sigma(X) = \sqrt{npq}$	$\xi_X(u) = (q + pe^{-2i\pi u})^n$

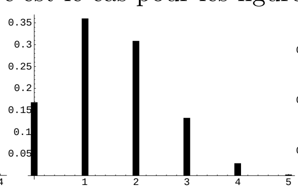
Remarque 3 *La formule donnant l'espérance semble assez naturelle. En effet, le nombre moyen de succès (qui correspond à la signification de l'espérance) est intuitivement égal au produit du nombre d'essais par la probabilité de réalisation d'un succès.*

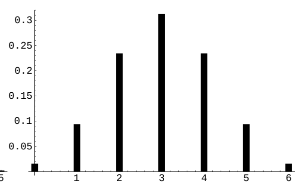
Propriétés de la distribution binomiale

Forme de la distribution binomiale

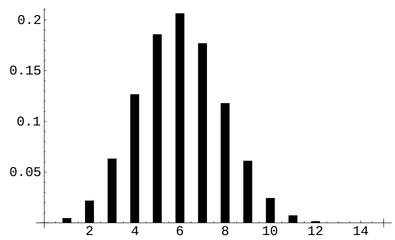
La représentation graphique de la distribution de la loi binomiale est habituellement présentée sous la forme d'un diagramme en bâtons. Puisque la loi dépend de n et p, nous aurons diverses représentations graphiques si nous faisons varier n et/ou p comme c'est le cas pour les figures suivantes.







$n = 4$, $p = 0.1$ $n = 5$, $p = 0.3$ $n = 6$, $p = 0.5$



$n = 15$, $p = 0.4$

On peut effectuer plusieurs remarques à propos de ces diagrammes.

a) La forme de la distribution est symétrique si $p = 1/2$, quelque soit n.

b) Elle est dissymétrique dans le cas où $p \neq 1/2$. Si p est inférieur à 0.50, les probabilités sont plus élevées du côté gauche de la distribution que du côté droit (asymétrie positive). Si p est supérieur à $1/2$, c'est l'inverse (asymétrie négative).

c) La distribution tend à devenir symétrique lorsque n est grand. De plus, si p n'est pas trop voisin de 0 ou 1, elle s'approchera de la distribution de la loi normale que l'on verra plus loin dans ce chapitre.

Somme de deux variables binomiales

Si X_1 et X_2 sont des variables *indépendantes* qui suivent des lois binomiales $B(n_1, p)$ et $B(n_2, p)$ respectivement, alors $X_1 + X_2$ suit une loi binomiale $B(n_1 + n_2, p)$.

Cette propriété s'interprète facilement : si X_1 représente le nombre de succès en n_1 épreuves identiques indépendantes et X_2 en n_2 épreuves indépendantes entre elles et indépendantes des premières avec la même probabilité de succès que les premières, alors $X_1 + X_2$ représente le nombre de succès en $n_1 + n_2$ épreuves identiques et indépendantes.

3.2 Distribution géométrique

3.2.1 Situation concrète

a) On effectue une épreuve de Bernoulli. Elle n'a donc que deux issues : le succès avec une probabilité p ou l'échec avec une probabilité $q = 1 - p$.

b) On répète l'épreuve jusqu'à l'apparition du premier succès.

c) Toutes les épreuves sont indépendantes entre elles.

Dans cette situation, on s'intéresse à la variable X = "nombre de fois qu'il faut répéter l'épreuve pour obtenir le premier succès".

Remarque 4 *On est donc dans les mêmes hypothèses que pour la loi binomiale, mais le nombre d'épreuves n'est pas fixé à l'avance. On s'arrête au premier succès.*

3.2.2 Distribution de probabilités

L'ensemble des valeurs prises par X est $1, 2, 3, \ldots$. On cherche la probabilité d'avoir recours à n épreuves pour obtenir le premier succès.

Ce succès a une probabilité de réalisation de p. Puisque c'est le premier, il a été précédé de $n-1$ échecs qui ont chacun eu la probabilité q de se produire. Étant donné l'indépendance des épreuves, on peut dire que la probabilité de réalisation de $n-1$ échecs suivis d'un succès est le produit des probabilités de réalisation de chacun des résultats,

$$p(X = n) = q^{n-1}p.$$

On dit que la variable aléatoire X suit une *loi géométrique de paramètre p*. On note $X \hookrightarrow G(p)$.

Remarque 5 *L'appellation* géométrique *vient du fait qu'en sommant toutes les probabilités, on obtient une série géométrique. En effet,*

$$\sum_{n=1}^{+\infty} q^{n-1}p = \frac{p}{1-q} = 1.$$

3.2.3 Paramètres descriptifs de la distribution

On calcule

$$\begin{aligned} \xi_X(u) &= \sum_{n=1}^{\infty} q^{n-1}pe^{-2i\pi un} \\ &= pe^{-2i\pi u}\sum_{k=0}^{\infty} q^k e^{-2i\pi uk} \\ &= \frac{pe^{-2i\pi u}}{1-qe^{-2i\pi u}}, \end{aligned}$$

et on en tire, en dérivant par rapport à u en $u = 0$, l'espérance et la variance.

$E(X) = 1/p$	$Var(X) = q/p^2$	$\sigma(X) = \sqrt{q}/p$	$\xi_X(u) = \frac{pe^{-2i\pi u}}{1-qe^{-2i\pi u}}$

Remarque 6 *On peut interpréter l'expression de l'espérance de façon intuitive. En effet en n épreuves, on s'attend à obtenir np succès et par conséquent, le nombre moyen d'épreuves entre deux succès devrait être* $\frac{n}{np} = \frac{1}{p}$.

3.2.4 Propriété remarquable de la distribution géométrique

La propriété la plus importante de la loi géométrique est sans doute d'être *sans mémoire*. En effet, la loi de probabilité du nombre d'épreuves à répéter jusqu'à l'obtention d'un premier succès dans une suite d'épreuves de Bernoulli identiques indépendantes est la même quel que soit le nombre d'échecs accumulés auparavant. Mathématiquement, cela se traduit par

$$p(X > n + m | X > n) = p(X > m).$$

On comprend intuitivement que cela découle de l'indépendance des épreuves qui sont toutes identiques. La loi géométrique est la seule distribution de probabilité discrète qui possède cette propriété. En effet, si une variable aléatoire Y à valeurs dans $\mathbf{N}$ satisfait, pour tous n, $m \in \mathbf{N}$, $p(Y > n + m|Y > n) = p(Y > m)$, alors $p(Y > m + n) = p(Y > n)p(Y > m)$. Posant $q = p(Y > 1)$, il vient $p(Y > n) = q^n$, d'où $p(Y = n) = p(Y > n - 1) - p(Y > n) = q^{n-1} - q^n = q^{n-1}(1 - q)$.

3.3 Distribution de Poisson

La loi de Poisson est attribuée à Simeon D. Poisson, mathématicien français (1781-1840). Cette loi fut proposée par Poisson dans un ouvrage qu'il publia en 1837 sous le titre "Recherche sur la probabilité de jugements en matière criminelle et en matière civile".

3.3.1 Situation concrète

Beaucoup de situations sont liées à l'étude de la réalisation d'un événement dans un intervalle de temps donné (arrivée de clients qui se présentent à un guichet d'une banque en une heure, apparitions de pannes d'un réseau informatique en une année, arrivée de malades aux urgences d'un hôpital en une nuit,....). Les phénomènes ainsi étudiés sont des phénomènes d'attente.

Pour décrire les réalisations dans le temps d'un événement donné, on peut

- soit chercher le nombre de réalisations de l'événement dans un intervalle de temps donné qui est distribué suivant une loi de Poisson.
- soit chercher le temps entre deux réalisations successives de l'événement qui est distribué suivant une loi exponentielle (voir section suivante).

On va voir que la première loi (loi de Poisson) peut être interprétée comme un cas limite d'une loi binomiale et la seconde comme un cas limite d'une loi géométrique.

3.3.2 Processus de Poisson

Précisons les hypothèses faites relativement à la réalisation de l'événement qui nous intéresse.

1. Les nombres de réalisations de l'événement au cours d'intervalles de temps disjoints sont des variables aléatoires indépendantes, c'est-à-dire que le nombre de réalisations au cours d'un intervalle de temps est indépendant du nombre de réalisations au cours d'intervalles de temps antérieurs.
2. La probabilité pour que l'événement se réalise une fois, au cours d'un petit intervalle de temps Δt, est proportionnelle à l'amplitude de l'intervalle et vaut $\alpha \Delta t$, où α est une valeur positive que l'on suppose constante tout au long de la période d'observation.
3. Il est très rare d'observer plus d'une fois l'événement au cours d'un petit intervalle de temps Δt, c'est-à-dire que la probabilité pour que l'événement se réalise plus d'une fois au cours de l'intervalle de temps Δt est négligeable.

Les hypothèses 1., 2., 3. caractérisent ce qu'on appelle un *processus de Poisson.* α est une constante du processus qui représente le nombre moyen de réalisations par unité de temps et que l'on appelle l'*intensité* du processus.

Sous ces hypothèses, la variable aléatoire X = "nombre de fois où l'événement considéré se réalise au cours d'un intervalle de temps de durée t" est distribuée suivant une loi de Poisson de paramètre $\lambda = \alpha t$.

3.3.3 Distribution de probabilités

Nous cherchons à déterminer la loi de probabilité de la variable X = "nombre de réalisations d'un événement donné pendant un intervalle de temps t", sachant que le nombre moyen de réalisations de cet événement par unité de temps est α. Or, nous connaissons déjà la loi de probabilités de la variable $Y =$ "nombre de réalisations d'un événement donné, de probabilité p, au cours de n essais". Il s'agit d'une loi binomiale $B(n, p)$.

Pour comprendre la relation entre ces deux lois, divisons l'intervalle de temps de longueur t, en n petits intervalles de temps disjoints de longueur $\Delta t = t/n$ pour n assez grand.

L'hypothèse 3. permet d'affirmer que dans chacun de ces n petits intervalles il n'y a principalement que deux possibilités : l'événement se réalise une fois ou ne se réalise pas (cela sera d'autant plus vrai que n est grand). Dans chaque intervalle, la variable "nombre de réalisations de l'événement" est une variable de Bernoulli.

L'hypothèse 2. permet d'affirmer que dans chacun de ces n petits intervalles, la probabilité de réalisation de l'événement est constante et égale à $\alpha\Delta t = \alpha t/n$. Les variables de Bernoulli ont donc toutes le même paramètre $p = \alpha t/n$.

L'hypothèse 1. permet d'affirmer que les n variables de Bernoulli sont indépendantes.

La somme de ces n variables de Bernoulli indépendantes de même paramètre $p = \alpha t/n$ est une variable qui suit la loi binomiale $B(n, \alpha t/n)$ et qui représente approximativement le nombre de réalisations de l'événement dans l'intervalle de temps t. Si on choisit n de plus en plus grand, on a de plus en plus d'intervalles, la probabilité de réalisations de l'événement dans chaque intervalle est de plus en plus petite et la distribution $B(n, \alpha t/n)$ se rapproche de plus en plus de la distribution que l'on cherche à déterminer, c'est-à-dire de la distribution de Poisson de paramètre αt. On conclut

Définition 7 *On peut considérer la loi de Poisson de paramètre λ comme la loi limite d'une loi binomiale $B(n, \lambda/n)$ lorsque n tend vers l'infini, le produit des paramètres $n.\lambda/n$ restant toujours constant égal à λ.*

On écrit $X \hookrightarrow P(\lambda)$.

Proposition 8 *La loi de Poisson de paramètre λ est donnée par*

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Preuve. Si Y suit une loi $B(n, \lambda/n)$, on sait que

$$\begin{aligned}
P(Y = k) &= \binom{k}{n}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k} \\
&= (1-\frac{\lambda}{n})^{n-k}\frac{\lambda^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k} \\
&= (1-\frac{\lambda}{n})^{n}\frac{\lambda^k}{k!}(1-\frac{\lambda}{n})^{-k}[\frac{n}{n}\times\frac{n-1}{n}\times\cdots\times\frac{n-k+1}{n}].
\end{aligned}$$

Chaque terme du produit entre crochets tend vers 1 lorsque n tend vers l'infini. Il y a k termes, c'est-à-dire un nombre fini. Donc le crochet tend vers 1. De même, $(1-\frac{\lambda}{n})^{-k}$ tend vers 1. De plus,

$$\ell n((1-\frac{\lambda}{n})^n) = n\ell n(1-\frac{\lambda}{n}) \sim n \times (-\frac{\lambda}{n})$$

tend vers $-\lambda$ lorsque n tend vers l'infini, donc $(1-\frac{\lambda}{n})^n$ tend vers $e^{-\lambda}$. On conclut que $P(Y = k)$ tend vers $e^{-\lambda}\lambda^k/k!$.

Remarque 9 *Il existe des tables donnant la fonction de densité et la fonction de répartition de la loi de Poisson en fonction des différentes valeurs de λ (pour $\lambda \leq 15$).*

3.3.4 Paramètres descriptifs de la distribution

On calcule, lorsque $X \hookrightarrow P(\lambda)$,

$$\xi_X(u) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}e^{-2i\pi uk} = e^{-\lambda}e^{\lambda e^{-2i\pi u}}.$$

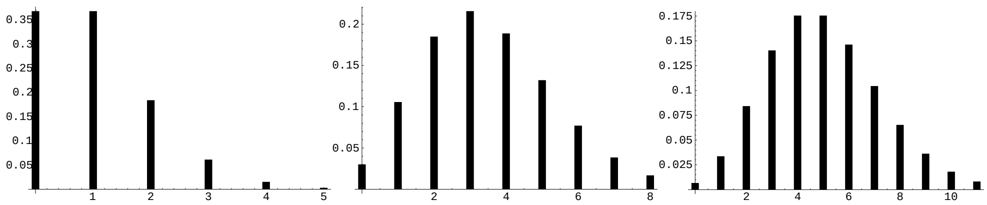
On en déduit le tableau

$E(X) = \lambda$	$Var(X) = \lambda$	$\sigma(X) = \sqrt{\lambda}$	$\xi_X(u) = e^{\lambda(e^{-2i\pi u}-1)}$

La loi $P(\lambda)$ est la loi limite de la loi $B(n, \lambda/n)$ lorsque n tend vers l'infini. On constate que l'espérance mathématique et la variance de la loi $B(n, \lambda/n)$ convergent vers celles de la loi $P(\lambda)$ lorsque n tend vers l'infini. Cela peut se vérifier directement, en appliquant le théorème de convergence dominée pour la mesure Peigne de Dirac (interversion d'une sommation et d'une limite).

3.3.5 Propriétés de la distribution de Poisson

Allure de la distribution



$\lambda = 1$ $\lambda = 3.5$ $\lambda = 5$

Commentaires.
- En général, le diagramme est dissymétrique par rapport à λ avec étalement vers la droite. Les valeurs élevées d'une variable de Poisson sont peu rencontrées.
- A mesure que λ augmente, la forme de la distribution tend à devenir symétrique et s'approche de celle de la loi normale que nous traiterons plus loin dans ce chapitre. Cela est vérifié pour $\lambda \geq 10$ et même acceptable pour $\lambda \geq 5$.

Approximation de la loi binomiale par la loi de Poisson

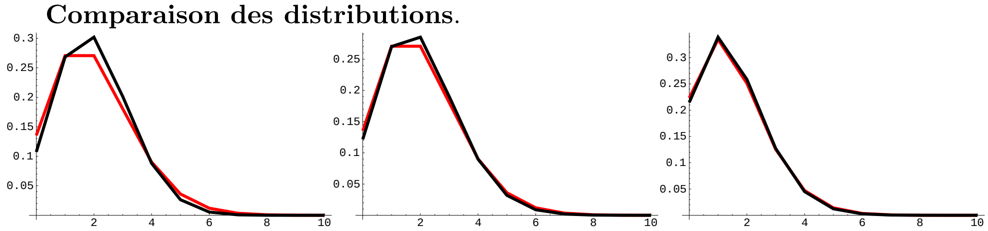
La loi binomiale dépend de deux paramètres n et p. Bien qu'il existe quelques tables, elle est n'est pas simple à utiliser. La loi de Poisson ne dépend que d'un paramètre ce qui la rend plus pratique. Il faut donc avoir toujours présent à l'esprit que, lorsque les conditions le permettent, on peut avoir intérêt à remplacer une loi binomiale par une loi de Poisson.

Lorsque n est grand et p petit, de telle façon que le produit $np = \lambda$ reste petit par rapport à n, la loi binomiale $B(n, p)$ peut être approchée par la loi de Poisson $P(\lambda)$ (revoir ce qui a été dit sur ce sujet dans le paragraphe "Distribution de probabilités"). Cette approximation s'appliquant lorsque p est petit, la loi de Poisson est appelée la loi des événements rares. En pratique, l'approximation est valable si $n > 20$, $p \leq 0.1$ et $np \leq 5$.

$\boxed{\text{On approche la loi } B(n, p) \text{ par la loi } P(np) \text{ dès que } n > 20,\ p \leq 0.1 \text{ et } np \leq 5.}$

RÈGLE IMPORTANTE. Lorsqu'on approche une loi par une autre, on choisit le ou les paramètres de la loi approchante de manière que l'espérance (et la variance lorsqu'on a suffisamment de paramètres) de la loi approchante soit égale à l'espérance (et la variance) de la loi approchée.

Comparaison des distributions.



En rouge	$B(10, 0.2)$	$B(20, 0.1)$	$B(30, 0.05)$
En noir	$P(2)$	$P(2)$	$P(1.5)$
	mauvaise approximation	approximation acceptable	bonne approximation

Somme de deux lois de Poisson

Si X_1 et X_2 sont des variables aléatoires *indépendantes* qui suivent des lois de Poisson de paramètres respectifs λ_1 et λ_2, alors $X_1 + X_2$ suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

3.3.6 Importance pratique de la loi de Poisson

Le physicien et le technologue rencontrent la loi de Poisson dans de nombreuses circonstances.

Le comptage en radioactivité, microanalyse X, analyse SIMS...

Un détecteur compte le nombre de particules qu'il reçoit pendant un temps de comptage t_0. Ce nombre X obéit à une loi de Poisson dont le paramètre λ est le produit du flux moyen de particules α pendant le temps de comptage t_0.

Conséquence. L'espérance mathématique de X est αt_0 et l'écart-type sur X est $\sqrt{\alpha t_0}$. Il en résulte que la variabilité de X (quotient de l'espérance sur l'écart-type) devient très grande quand le nombre moyen de particules observé est petit. Si les valeurs prises par X sont proches de 10 correspondant à une espérance αt_0 proche de 10, l'écart-type est proche de 3, ce qui donne une variabilité de

30%. Ceci est parfaitement visible sur le "profil SIMS[1]" présenté ci-dessous. Le bruit devient important lorsque X est entre 10 et 15.

PROFIL S.I.M.S.

Dans la mesure du possible, l'expérimentateur sera alors amené à augmenter le temps de comptage pour augmenter les valeurs de X.

Contrôle de qualité

On effectue un contrôle de qualité sur des composants fabriqués par une unité de production. Un composant est réputé bon (ou mauvais) lorsqu'il satisfait (ou non) à des spécifications. Le critère de choix est donc qualitatif. Le taux moyen de rebuts (pièces mauvaises) est appelé p. On effectue un tirage de n pièces. La probabilité de trouver k pièces mauvaises dans cet échantillon de taille n est donnée par une loi de Poisson $P(np)$, car on peut approcher la loi binomiale par la loi de Poisson (*loi des événements rares*).

Lorsque np est petit, le rapport devient grand et la variabilité sur k rend le contrôle imprécis. Ceci explique que, dans un contrôle de qualité, la taille des échantillons tirés de la population des pièces fabriquées doit être au moins de l'ordre de 100.

Dénombrement d'événements survenant dans un espace délimité

La loi de Poisson permet de modéliser aussi bien le nombre d'événements survenant pendant un temps donné que le nombre d'événements survenant dans un espace délimité.

Par exemple, si on appelle X le nombre de particules bombardant une cible de surface S soumise à une irradiation de fluence F (mesurée en m^{-2}), alors X suit une loi $P(FS)$. FS est donc analogue au αt_0 du comptage en temps.

La loi de Poisson sert donc aussi à décrire des phénomènes de localisation spatiale et non plus seulement temporelle, c'est-à-dire qu'elle modélisera aussi bien le nombre d'accidents qui peuvent survenir en une matinée que le nombre d'accidents qui peuvent survenir sur une section donnée d'autoroute.

3.4 Distribution exponentielle

3.4.1 Situation concrète

On se place dans le cas d'un phénomène d'attente décrit au paragraphe 3 et on s'intéresse à la variable aléatoire qui représente le temps d'attente pour la réalisation d'un événement ou le temps d'attente entre la réalisation de deux événements successifs. Si on se place dans le cas où l'intensité α du processus de Poisson est constante, ce temps d'attente suit une loi exponentielle de paramètre α.

Exemple. Lorsque l'événement attendu est la mort d'un individu (ou la panne d'un équipement), α s'appelle le taux de mortalité (ou le taux de panne).

[1] S.I.M.S. (Secondary Ions Mass Spectroscopy) : méthode permettant une analyse quantitative des impuretés dans une masse solide. Un faisceau d'ions primaires abrase le matériau. Parmi les atomes arrachés à la cible, certains sont ionisés (ions secondaires). Ils sont triés en masse au moyen d'une déflexion magnétique et comptés pendant des intervalles de temps t_0 au cours du processus d'abrasion. On obtient ainsi le profil de concentration de l'impureté dans le matériau.

Dire qu'il a une valeur constante, c'est supposer qu'il n'y a pas de vieillissement (ou pas d'usure s'il s'agit d'un équipement), la mort ou la panne intervenant de façon purement accidentelle.

3.4.2 Distribution de probabilité

On veut déterminer la loi de la variable T = "temps d'attente entre la réalisation de deux événements successifs" où le nombre moyen de réalisations de l'événement par unité de temps est α. Pour cela, nous allons procéder comme dans le paragraphe 3 : Si t est la longueur de l'intervalle de temps sur lequel dure notre étude, nous le divisons en n petits intervalles de longueur t/n. Appelons X la variable aléatoire représentant le nombre d'intervalles de temps que l'on doit laisser s'écouler pour obtenir la réalisation de l'événement suivant. Chaque intervalle possédant la même probabilité $\alpha t/n$ de voir l'événement se produire, X suit, par définition, la loi géométrique de paramètre $\alpha t/n$. Le temps d'attente T est alors le produit de ce nombre d'intervalles par le temps de chaque intervalle, $T = \alpha t/n \times X$.

On cherche $p(T > t_0) = p(X > nt_0/t)$, et lorsque n tend vers l'infini on obtient

$$p(T > t_0) = e^{-\alpha t_0} = \int_{t_0}^{+\infty} \alpha e^{-\alpha u}\, du.$$

Ceci nous permet d'affirmer que la fonction de densité de la variable T est $f(x) = \alpha e^{-\alpha x}$ si $x > 0$, 0 sinon.

Définition 10 *La loi exponentielle de paramètre α décrit la distribution d'une variable continue X qui ne prend que des valeurs positives selon la fonction de densité $f(x) = \alpha e^{-\alpha x}$. On la note $Exp(\alpha)$.*

3.4.3 Paramètre descriptifs de la distribution

On vient de voir que la loi exponentielle est la loi limite d'une loi géométrique. On a $T \sim \frac{t}{n}X$ où X suit la loi géométrique de paramètre $\frac{\alpha t}{n}$. Or $E(T) \sim \frac{t}{n}E(X) = \frac{t}{n}(\frac{\alpha t}{n})^{-1} = \frac{1}{\alpha}$ et $Var(T) \sim \frac{t^2}{n^2}Var(X) = \frac{t^2}{n^2}\frac{1-\frac{\alpha t}{n}}{(\frac{\alpha t}{n})^2} \sim \frac{1}{\alpha^2}$.

Proposition 11

$$E(T) = \frac{1}{\alpha}, \quad Var(T) = \frac{1}{\alpha^2}.$$

Remarque 12 *On peut très facilement retrouver ces résultats en effectuant un calcul direct à partir de la fonction de densité en utilisant les formules de définition de l'espérance et la variance (peut-être pourriez-vous le faire à titre d'exercice')*

3.4.4 Propriétés de la distribution exponentielle

1. Comme la loi géométrique, la loi exponentielle est sans mémoire. C'est la seule loi continue qui possède cette propriété. Elle provient bien entendu du fait que le paramètre α est constant.

2. Une somme de n variables indépendantes de même loi exponentielle de paramètre α n'est pas une variable de loi exponentielle mais une variable qui suit une loi gamma de paramètres n et α. Une telle loi est aussi appelée loi d'ERLANG d'ordre n. Elle représente le temps d'attente requis avant qu'un événement se réalise n fois .
3. La loi exponentielle est aussi couramment utilisée dans les problèmes de datation en géochronologie.

3.5 Distribution normale

Du point de vue historique, la nature et l'importance exceptionnelle de cette loi furent pressenties en 1773 par Abraham de Moivre lorsqu'il considéra la forme limite de la loi binomiale.

En 1772, Simon Laplace l'étudia dans sa théorie des erreurs. Mais c'est seulement en 1809 pour Carl Friedrich Gauss et en 1812 pour Simon Laplace qu'elle prit sa forme définitive. C'est ainsi qu'on l'appelle tantôt loi de Laplace, tantôt loi de Gauss, tantôt loi de Laplace-Gauss. On trouve aussi l'expression, consacrée par une longue tradition, de loi normale (ce qui ne signifie pas pour autant que les autres lois soient « anormales »). Elle jouit d'une importance fondamentale car un grand nombre de méthodes statistiques reposent sur elle. Ceci est lié au fait qu'elle intervient comme loi limite dans des conditions très générales.

Pour faire ressortir toute son importance et sa forme, W.J. Youden, du National Bureau of Standards, a eu l'ingénieuse idée de la présenter telle qu'elle apparaît ci-dessous.

La
loi normale
des erreurs
constitue l'une
des généralisations
les plus étendues de
la philosophie naturelle
dans l'histoire de l'humanité.
Elle est un outil précieux pour la
recherche en sciences physiques et
sociales ainsi qu'en médecine, en agriculture
et en génie. Elle est indispensable à l'analyse et à
l'interprétation des données obtenues par l'observation ou
l'expérience.

3.5.1 Situation concrète

On rencontre souvent des phénomènes complexes qui sont le résultat de causes nombreuses, d'effet faible, et plus ou moins indépendantes. Un exemple typique est celui de l'erreur commise sur la mesure d'une grandeur physique. Cette erreur résulte d'un grand nombre de facteurs tels que : variations incontrôlables de la température ou de la pression, turbulence atmosphérique, vibrations de l'appareil de mesure, etc...Chacun des facteurs a un effet faible, mais l'erreur résultante peut ne pas être négligeable. Deux mesures faites dans des conditions que l'expérimentateur considère comme identiques pourront alors donner des résultats différents.

Donc dès que nous seront dans une situation où la distribution dépend de causes
- en grand nombre et indépendantes,
- dont les effets s'additionnent,
- dont aucune n'est prépondérante,

alors nous serons en présence de la distribution normale. C'est le cas, par exemple :
- En métrologie, pour la distribution des erreurs d'observation.
- En météorologie, pour la distribution de phénomènes aléatoires tels que la température et la pression.
- En biologie, pour la distribution de caractères biométriques comme la taille ou le poids d'individus appartenant à une population homogène. En technologie, pour la distribution des cotes des pièces usinées.
- En économie, pour les fluctuations accidentelles d'une grandeur économique (production, ventes, ...) autour de sa tendance, etc.....

3.5.2 Distribution de probabilité

Définition 13 *Une variable aléatoire continue suit une loi normale si l'expression de sa fonction de densité de probabilités est de la forme :*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m}{\sigma})^2}, \quad x \in \mathbf{R}.$$

La loi dépend des deux réels m et σ appelés paramètres de la loi normale. On la note $\mathcal{N}(m, \sigma)$.

Remarque 14
1. *Une fonction de densité de probabilité étant toujours positive, le paramètre σ est donc un réel strictement positif.*
2. *On démontre que f est bien une fonction de densité de probabilité car $\int_{\mathbf{R}} f(x)\,dx = 1$. Pour le démontrer on utilise que $\int_{\mathbf{R}} e^{-x^2/2}\,dx = \sqrt{2\pi}$ (c'est l'intégrale de Gauss).*
3. *La loi normale étant tabulée, cette expression nous sera de peu d'utilité. Il est important néanmoins de préciser à quoi correspondent m et σ.*

3.5.3 Paramètre descriptifs de la distribution

La fonction caractéristique d'une variable normale standard X vaut

$$\xi_X(u) = e^{-2i\pi m u - 2\pi^2\sigma^2 u^2}.$$

On en déduit, à l'aide de la formule qui exprime espérance et variance à partir des dérivées de la fonction caractéristique, que

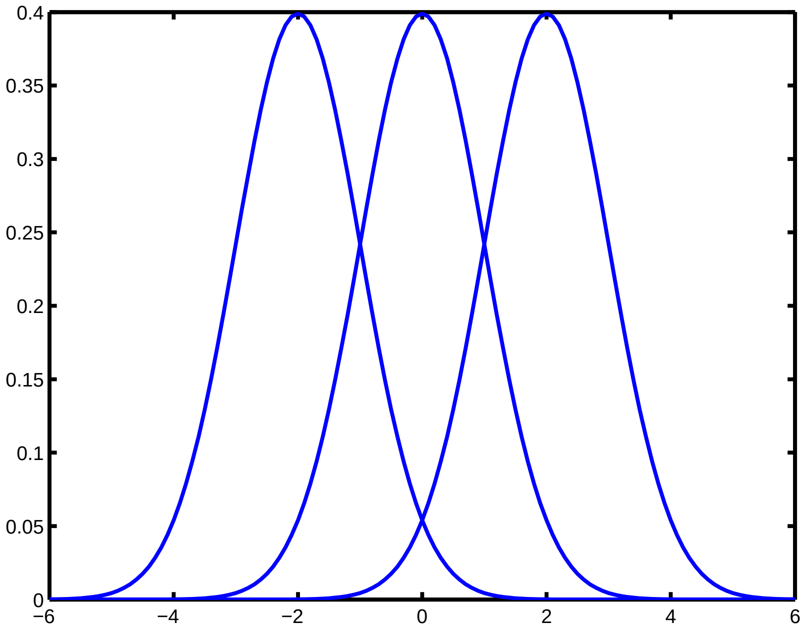
$$E(X) = m, \quad Var(X) = \sigma^2, \quad \sigma(X) = \sigma.$$

On peut aussi faire le calcul directement, à partir de l'intégrale de Gauss.

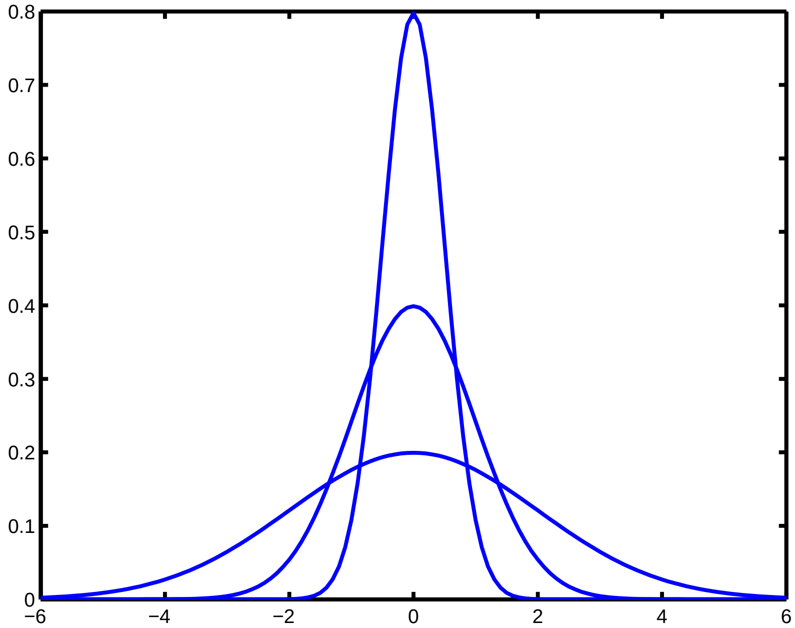
3.5.4 Propriétés de la distribution normale

Forme de la distribution normale

La fonction de densité de probabilités de la loi normale a la forme d'une « courbe en cloche ». En fait il ne s'agit pas d'une courbe unique mais plutôt d'une famille de courbes dépendant de m et σ.



Ecarts types identiques, espérances -2, 0, 2 différentes



Espérances identiques, écarts types 0.5, 1, 2 différents

On peut effectuer quelques remarques à propos de ces courbes.

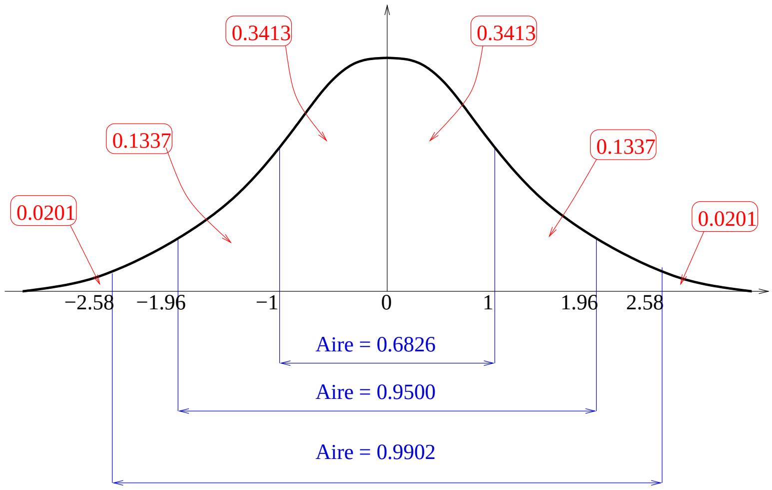
a) La distribution est symétrique par rapport à la droite d'équation $x = m$. Donc l'aire sous la courbe de part et d'autre de cette droite est égale à 0.5.

b) La distribution est d'autant plus étalée que σ est grand.

c) L'axe des abscisses est une asymptote et l'aire sous la courbe à l'extérieur de l'intervalle$[m - 3\sigma, m + 3\sigma]$ est négligeable. Pour fixer les idées, on peut indiquer que

$$\begin{aligned} p(m - \sigma < X < m + \sigma) &= 0.6826 \\ p(m - 2\sigma < X < m + 2\sigma) &= 0.9544 \\ p(m - 3\sigma < X < m + 3\sigma) &= 0.9974. \end{aligned}$$

Cela peut être visualisé sur le graphique ci-après.



Loi $\mathcal{N}(0,1)$.

d) σ représente la différence des abscisses entre le sommet de la courbe et le point d'inflexion.

e) La longueur à mi-hauteur de la courbe (L.M.H. ou en anglais F.W.H.M. Full Width Half Maximum) vaut 2.35σ. Cette distance est souvent employée par le spectroscopiste pour déterminer expérimentalement σ. Cette méthode doit cependant être utilisée avec précaution car il faut s'assurer que les "bruits" permettent d'observer correctement le "pied" de la courbe.

Somme de deux variables normales

Soient X_1 et X_2 deux variables indépendantes. Si X_1 suit $\mathcal{N}(m_1, \sigma_1)$ et X_2 suit $\mathcal{N}(m_2, \sigma_2)$, alors $X_1 + X_2$ suit $\mathcal{N}(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

Loi normale centrée réduite ou loi normale standardisée

Nous avons vu dans le chapitre 2 qu'à toute variable aléatoire X ,on pouvait associer une variable dite standardisée $\dfrac{X - E(X)}{\sigma(X)}$ d'espérance nulle et de variance unité (ceci résultait des propriétés de translation et de changement d'échelle). On montre assez facilement que si on effectue cette transformation sur une variable suivant une loi normale, la variable standardisée suit encore une loi normale mais cette fois-ci de paramètres 0 et 1. La loi standardisée est appelée loi normale centrée réduite, et notée $\mathcal{N}(0,1)$. Donc si X suit $\mathcal{N}(m, \sigma)$, on pose $T = \dfrac{X - m}{\sigma}$ et T suit $\mathcal{N}(0,1)$.

On peut résumer la correspondance de la façon suivante :

$X \rightharpoonup \mathcal{N}(m, \sigma)$ $E(X) = m$ $Var(X) = \sigma^2$	$T = \frac{X-m}{\sigma}$	$T \rightharpoonup \mathcal{N}(0,1)$ $E(T) = 0$ $Var(T) = 1$

Il faut garder à l'esprit que concrètement T est le nombre d'écarts-type entre la valeur de X et la moyenne.

La loi $\mathcal{N}(0,1)$ est tabulée à l'aide la fonction de répartition des valeurs positives. Elle donne les valeurs de $\Phi(t) = p(0 \leq T \leq t) = \int_0^t \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$ pour $t > 0$. Ce nombre représente l'aire sous la courbe représentative de la distribution et au dessus de l'intervalle $[0,t]$. Pour cette raison la table de la loi normale est aussi appelée table d'aires. Cette table ne dépend d'aucun paramètre, mais permet cependant de déterminer les probabilités de n'importe quelle distribution normale !

Comment utiliser la table d'aires ?

La première colonne de la table indique les unités et les dixièmes des valeurs de T alors que les centièmes des valeurs de T se lisent sur la ligne supérieure de la table. La valeur trouvée à l'intersection de la ligne et de la colonne adéquates donne l'aire cherchée.

a) Je cherche la valeur de A l'intersection de la ligne "0.5" et de la colonne "0.00", je lis 0.1915.

b) Je cherche la valeur de $p(-0.5 \leq T \leq 0)$. J'utilise la symétrie de la courbe par rapport à l'axe des ordonnées et j'en conclus que $p(-0.5 \leq T \leq 0) = p(0 \leq T \leq 0.5) = 0.1915$. Et que pensez-vous de la valeur de $p(-0.5 < T < 0)$?

c) Je cherche la valeur de $p(-2.24 \leq T \leq 1.12)$. L'aire cherchée correspond à la somme suivante

$$p(-2.24 \leq T \leq 1.12) = p(-2.24 \leq T \leq 0) + p(0 < T \leq 1.12) = 0.4875 + 0.3686 = 0.8561.$$

d) Je cherche la valeur de $p(1 \leq T \leq 2)$. L'aire cherchée correspond à la différence suivante

$$p(1 \leq T \leq 2) = p(0 \leq T \leq 2) - p(0 \leq T \leq 1) = 0.4772 - 0.3413 = 0.1359.$$

e) Je cherche la valeur t de T telle que $p(0 \leq T \leq t) = 0.4750$. C'est le problème inverse de celui des exemples précédents. Il s'agit de localiser dans la table l'aire donnée et de déterminer la valeur de T correspondante. Je trouve $t = 1.96$.

Remarque 15 *Si la valeur de l'aire ne peut se lire directement dans les valeurs de la table, on pourra toujours effectuer une interpolation linéaire entre deux valeurs adjacentes ou prendre la valeur la plus proche.*

3.6 Approximation par des lois normales

3.6.1 Théorème central limite (ou de tendance normale)

Théorème 1 *Hypothèses : Soit une suite de variables aléatoires $X_1, X_2, \ldots, X_n$ vérifiant les conditions suivantes :*

1. *Les variables sont indépendantes.*
2. *Leurs espérances mathématiques $m_1, m_2, \ldots, m_n$ et leurs variances $Var(X_1)$, $Var(X_2)$,... $Var(X_n)$ existent toutes.*
3. $\lim_{n\to\infty} \dfrac{\sup_k Var(X_k)}{\sum_{j=1}^n Var(X_j)} = 0.$

Conclusion : La distribution de la variable somme $X = X_1 + X_2 + \cdots + X_n$ se rapproche de la distribution normale lorsque n tend vers l'infini.

Idée de la preuve, dans le cas particulier où les variables X_k ont toutes la même loi.

Quitte à standardiser, on peut supposer que X_k est d'espérance nulle et de variance 1. Alors sa fonction caractéristique a un développement limité, au voisinage de $u = 0$, de la forme

$$\xi_{X_k}(u) = 1 - 2\pi^2 u^2 + \cdots$$

Posons

$$Y = \frac{1}{\sqrt{n}} X = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

Alors

$$\begin{aligned}
\xi_Y(u) &= \xi_X\left(\frac{u}{\sqrt{n}}\right) \\
&= \xi_{X_k}\left(\frac{u}{\sqrt{n}}\right)^n \\
&= \left(1 - \frac{2\pi^2 u^2}{n} + \cdots\right)^n \\
&\sim e^{-2\pi^2 u^2},
\end{aligned}$$

qui est la fonction caractéristique de la distribution normale standard $\mathcal{N}(0, 1)$.

Remarque 16 *C'est ce théorème très important qui nous permet d'affirmer que la situation concrète énoncée au début de ce paragraphe nous met en présence d'une loi normale. En effet,*

- *$X_1, X_2, \ldots, X_n$ correspondent aux différents facteurs de fluctuations.*
- *Le grand nombre de causes est assuré par le fait que n tend vers l'infini.*
- *L'indépendance parle d'elle-même.*
- *Additionner les effets revient à considérer la variable somme.*
- *Dire qu'aucun facteur n'est prépondérant est traduit par l'hypothèse (3) du théorème.*

En pratique, ceci se vérifie dès que $n \geq 30$.

3.6.2 Approximation de la loi binomiale par la loi normale

Une variable qui suit une loi binomiale $\mathcal{B}(n, p)$ peut toujours être considérée comme une somme de n variables de Bernoulli indépendantes de même paramètre p.

$$X = X_1 + \cdots + X_n,$$

où X_i sont des variables de Bernoulli. Les hypothèses du théorème centrale limite étant vérifiées, on peut affirmer que, lorsque n tend vers l'infini, la loi binomiale $\mathcal{B}(n, p)$ tend vers une loi normale. La loi normale qui l'approche le mieux est celle qui possède la même espérance np et le même écart-type $\sqrt{npq}$, $q = 1 - p$.

Or la distribution binomiale est asymétrique sauf lorsque $p = 1/2$. La distribution normale, elle, est symétrique. L'approximation sera valable lorsque p n'est pas trop voisin de 0 ou 1 et sera d'autant meilleure que p est proche de $1/2$ et que n est grand.

En pratique :

$$\boxed{\text{On approche la loi } \mathcal{B}(n,p) \text{ par la loi } \mathcal{N}(np, \sqrt{npq}) \text{ dès que } \begin{cases} n \geq 30 \\ np \geq 15 \\ nq \geq 15 \end{cases}}$$

3.6.3 La correction de continuïté

Cette approximation pose deux problèmes.

1. *On remplace une distribution concernant un nombre fini de valeurs par une distribution sur* **R** *tout entier.*

   Étant donné qu'à l'extérieur de l'intervalle $[m-3\sigma, m+3\sigma]$ la distribution normale est presque nulle, cela ne pose pas de problèmes.

2. *On remplace une distribution discrète par une distribution continue.*

   Il nous faut donc appliquer ce qu'on appelle une *correction de continuité*. Si on nomme X la variable binomiale et Y la variable normale, on remplacera une valeur k de X par un intervalle de Y centré sur k et d'amplitude 1, ce qui signifie que l'on écrit

   $$p(X = k) \simeq p(k - \frac{1}{2} < Y < k + \frac{1}{2}).$$

   Dans la pratique lorsque n est très grand, cette correction n'est pas nécessaire. On l'effectuera cependant si on souhaite une grande précision.

Remarque 17 *Remplacer une loi binomiale par une loi normale simplifie considérablement les calculs.*

En effet les tables de la loi binomiale dépendent de deux paramètres et les valeurs de n dans ces tables sont limitées supérieurement par 20. La loi normale, elle, après standardisation ne dépend d'aucun paramètre .

3.6.4 Approximation de la loi de Poisson par la loi normale

On démontre qu'on peut aussi approcher la loi de Poisson par la loi normale pour les grandes valeurs du paramètre de la loi de Poisson. La seule qui puisse convenir est celle qui a même espérance et même variance. On approche donc la loi $\mathcal{P}(\lambda)$ par la loi $\mathcal{N}(\lambda, \sqrt{\lambda})$. En pratique, cela s'applique dès que $\lambda \geq 16$.

$$\boxed{\text{On approche la loi } \mathcal{P}(\lambda) \text{ par la loi } \mathcal{N}(\lambda, \sqrt{\lambda}) \text{ dès que } \lambda \geq 16}$$

Remarque 18 *La loi de Poisson étant elle aussi une loi discrète, on peut avoir à appliquer la correction de continuité.*

3.7 Quelques conseils pour résoudre les problèmes

Voici, lorsqu'elle s'applique, une méthode de travail qui peut guider votre démarche.

1. Suite à l'énoncé du problème, identifier correctement à l'aide de mots la variable aléatoire que vous allez considérer.
2. Préciser les valeurs possibles que peut prendre cette variable.
3. Identifier correctement la loi de probabilité qu'elle suit en essayant de reconnaître dans le problème une situation type.
4. Déterminer les paramètres de la loi.
5. Utiliser les formules théoriques ou les tables pour déterminer les probabilités demandées. Face à de longs calculs et en l'absence de tables correspondant à vos ou votre paramètre, penser à approcher votre loi par une autre.

3.7.1 Quelques exercices types

Exercice 19 *Supposons qu'une tentative pour obtenir une communication téléphonique échoue (par exemple, parce que la ligne est occupée) avec la probabilité 0.25 et réussisse avec la probabilité 0.75. On suppose que les tentatives sont indépendantes les unes des autres. Quelle est la probabilité d'obtenir la communication si l'on peut effectuer trois tentatives au maximum ?*

Solution de l'exercice 19. *3 essais.*

Nous nous intéressons à la variable $X =$ « nombre de tentatives nécessaires pour obtenir la communication », ce que l'on peut considérer comme le nombre d'essais à faire pour obtenir le premier succès. X suit une loi géométrique de paramètre $p = 0.75$.

On cherche à déterminer $p(X \leq 3) = p(X = 1) + p(X = 2) + p(X = 3)$.

- On peut obtenir la communication au 1er essai. On a pour cela une probabilité $p(X = 1) = 0.75$.
- On peut obtenir la communication au 2ème essai. On a pour cela une probabilité $p(X = 2) = 0.25 \times 0.75 = 0.1875$.
- On peut obtenir la communication au 3ème essai. On a pour cela une probabilité $p(X = 3) = 0.252 \times 0.75 = 0.0469$.

Finalement la probabilité d'obtenir la communication en trois essais maximum est $0.75 + 0.1875 + 0.0469 = 0.9844$ soit 98.5 %.

Exercice 20 *Un fabricant de pièces de machine prétend qu'au plus 10% de ses pièces sont défectueuses. Un acheteur a besoin de 120 pièces. Pour disposer d'un nombre suffisant de bonnes pièces, il en commande 140. Si l'affirmation du fabricant est valable, quelle est la probabilité que l'acheteur reçoive au moins 120 bonnes pièces ?*

Solution de l'exercice 20. *Bonnes pièces.*

Appelons X la variable aléatoire correspondant au "nombre de bonnes pièces dans le lot de 140 pièces".

X prend ses valeurs entre 0 et 140. De plus pour chaque pièce, on n'a que deux éventualités : elle est bonne ou elle est défectueuse. La probabilité qu'une pièce soit défectueuse est 0.1. Par conséquent elle est bonne avec la probabilité 0.9. On est donc dans une situation type : X suit la loi binomiale $\mathcal{B}(140, 0.9)$ de paramètres $n = 140$ et $p = 0.9$.

On veut déterminer la probabilité que l'acheteur reçoive au moins 120 bonnes pièces sur les 140, soit $X \geq 120$. A priori, il nous faudrait calculer la somme des probabilités $p(X = 120) + p(X = 121) + \cdots + p(X = 140)$, ce qui serait épouvantablement long. On approxime donc la loi binomiale par une loi tabulée.

Comme $n \geq 30$, $np = 126 \geq 15$ et $nq = 14$, on pourra approcher la loi binomiale par une loi normale. On choisit la loi normale qui a la même espérance et le même écart-type. Donc X qui suit la loi $\mathcal{B}(140, 0.9)$ sera approchée par Y qui suit la loi $\mathcal{N}(126, 3.55)$.

Pour remplacer une loi discrète par une loi continue, il est préférable d'utiliser la correction de continuité,

$$p(X \geq 120) \simeq p(Y > 119.5).$$

On se ramène enfin à la loi normale centrée réduite. On pose $T = \frac{Y-126}{3.55}$, et

$$\begin{aligned} p(Y > 119.5) &= p(T > \frac{119.5 - 126}{3.55}) = p(T > -1.83) \\ &= p(T < 1.83) = 0.5 + \Phi(1.83) = 0.97. \end{aligned}$$

Conclusion : l'acheteur a 97 chances sur 100 de recevoir 120 bonnes pièces sur les 140 achetées.

Exercice 21 *Les statistiques antérieures d'une compagnie d'assurances permettent de prévoir qu'elle recevra en moyenne 300 réclamations durant l'année en cours. Quelle est la probabilité que la compagnie reçoive plus de 350 réclamations pendant l'année en cours ?*

Solution de l'exercice 21. *Réclamations.*

La variable X qui nous intéresse est le "nombre de réclamations reçues pendant une année". Il s'agit du nombre de réalisations d'un événement pendant un intervalle de temps donné. X suit donc une loi de Poisson. Le nombre moyen de réalisations dans une année est 300. Cette valeur moyenne est aussi le paramètre de la loi de Poisson. Donc X suit la loi $\mathcal{P}(300)$.

On cherche à déterminer $p(X > 350)$. Il n'y a pas de table de la loi de Poisson pour cette valeur du paramètre. Il nous faut donc approcher X qui suit la loi de Poisson $\mathcal{P}(300)$ par Y qui suit la loi normale de même espérance et de même écart-type, c'est-à-dire $\mathcal{N}(300, \sqrt{300})$.

Ici aussi, on remplace une loi discrète par une loi continue. Il faut donc appliquer la correction de continuité

$$p(X > 350) = p(X \geq 351) \simeq p(Y > 350.5).$$

On se ramène finalement à la loi normale centrée réduite. On pose $T = \frac{Y-300}{\sqrt{300}}$.

$$p(Y > 350.5) = p(T > \frac{350.5 - 300}{\sqrt{300}}) = p(T > 2.92) = 0.5 - \Phi(2.92) = 0.0017.$$

La compagnie d'assurances a donc 0.17% de chances de recevoir plus de 350 réclamations en un an.

Exercice 22 *Le nombre moyen de clients qui se présentent à la caisse d'un supermarché sur un intervalle de 5 minutes est de 10. Quelle est la probabilité qu'aucun client ne se présente à la caisse dans un intervalle de deux minutes (deux méthodes possibles) ?*

Solution de l'exercice 22. *Solution n^0 1.*

Considérons la variable aléatoire $X =$ "nombre de clients se présentant à la caisse dans un intervalle de deux minutes". Nous reconnaissons une situation type et la variable X suit une loi de Poisson. Vu qu'en moyenne 10 clients se présentent en 5 mn, l'intensité α du processus est de 2 clients par minute, $\alpha = 2$. Or le paramètre de la loi de Poisson est αt_0, t_0 étant ici 2 minutes. D'où $\lambda = 4$.

On cherche à calculer $p(X = 0)$. D'après la formule du cours, $p(X = 0) = e^{-\lambda} = e^{-4} = 0.018$.

Solution de l'exercice 22. *Solution n^0 2.*

Considérons à présent la question sous un autre angle en s'intéressant au temps d'attente Y entre deux clients. Le cours nous dit que la loi suivie par une telle variable est une loi exponentielle. Son paramètre α est l'intensité du processus de Poisson soit ici $\alpha = 2$. Y suit donc la loi $Exp(2)$.

Sa fonction de densité est $2e^{-2x}$ pour $x > 0$ exprimé en minutes. On en déduit que

$$p(Y \geq 2) = \int_2^{+\infty} 2e^{-2x}\, dx = [-e^{-2x}]_2^{+\infty} = e^{-4} = 0.018.$$